ADVANCING GENOMICS AND BIOINFORMATICS

NOVOGENE is teaming up with laboratories worldwide for
LIFE SCIENCES AND HUMAN HEALTH RESEARCH.

Novogene's engineers prepare libraries for genomic sequencing with an integrated data and workflow management.

Testing reagents for clinical diagnosis.

The manual operation laboratory at Novogene Sequencing Center.

Driven by an unprecedented volume of data produced from high-throughput sequencing technologies, multiomics data analytics have enabled comprehensive studies of complex biological processes underlying agricultural breeding and cancer treatment.

Ruiqiang Li, a bioinformatician, and CEO of Novogene, who founded the company in Beijing in 2011, explains that the rapid growth of sequencing throughput, and cost reduction, have made next-generation sequencing (NGS) more accessible to support microbial, plant and animal studies, as well as disease diagnostics, and precision medicine research.

"We have cooperated with more than 5,600 universities, research institutes, hospitals and enterprises, in more than 70 countries across six continents," says Li.

Li's global teams and laboratories for genomic sequencing and bioinformatics are stepping up to support integrated data and workflow management. Along with Zhiying Ma from Hebei Agricultural University in China, Novogene's team, led by chief scientist, Shilin Tian, has presented findings and strategies to identify genomic variation and loci influencing cotton fibre quality and yield, summarized in two papers published in 2018 and 2021 in *Nature Genetics*.

By studying the associations between genotype and phenotype in natural cotton populations, they hope to identify allelic variations and candidate genes, as well as targets for molecular selection and genetic manipulation for better breeds.

Their quest zoomed in on upland cotton (*Gossypium hirsutum*), a species accounting for cultivars in 90% of global cotton production. Their 2018 studies included resequencing a core collection of 419 sample groups, comprising genetically diverse and locally adapted populations, known as landraces, as well as improved cultivars. They also performed phenotyping across 12 environments and conducted genome-wide association studies of 13 fibre-related traits based on a single nucleotide polymorphisms (SNP) set.

Efforts to introduce beneficial traits of modern cotton cultivars remain limited by an inadequate understanding of their genomic basis, while the genetic effects of the large-scale structural variations underlying agronomic traits also remain unclear.

"Our 2021 study went on to generate two genomes of these two modern cotton cultivars," says Tian. "Our first challenge was to assemble an accurate reference genome, which we worked to refine to complete the gaps in these cotton genomes."

Another achievement is the identification of the large-scale structural variations influencing agronomic traits. Among 446 significantly associated structural variations, those encouraging better yield were found to be mainly located in At subgenome, while the incidence of structural variations was higher in Dt subgenome. The latter indicates stronger selection during species formation and variety development, giving better fibre quality and wilt resistance.

"Our work stimulates new ideas in gene sequencing and bioinformatics research and informs studies into improving the commercial breeding of other plants," he says.

NEW CLINICAL POSSIBILITIES
Caused by the accumulation of somatic mutations, cancer leads to the formation of distinct populations of cells, called clones, which are the main cause of relapse and resistance to treatment. Somatic mutation analysis, a standard oncological practice to identify mutations sensitive or resistant to therapies, allows selection of targeted therapies built upon an individual's tumour profile.

Novogene's research collaboration with international academia extends to evolutionary genomics and prediction of anti-cancer drug sensitivity for patients, such as mesothelioma, a cancer caused by asbestos.
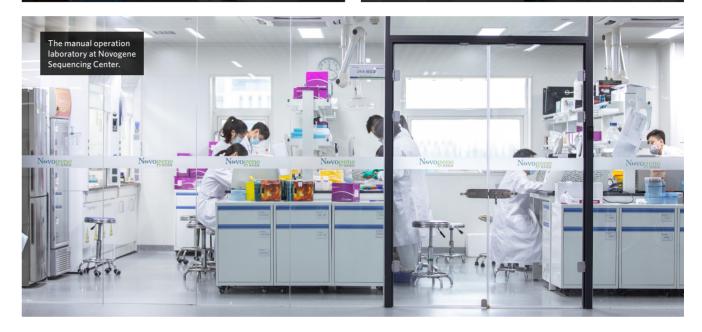
"Malignant pleural mesothelioma (MPM) is a rare, incurable cancer that exhibits links to exposure to asbestos.The United Kingdom has the highest incidence of MPM in the world," says Dean Fennell, the chair of thoracic medical oncology at Leicester University.

Mesothelioma takes on average 30 years to form in a patient's body. However, the evolutionary pathway remains mostly unknown. Understanding how MPM evolves could help us identify and prioritize clonal drug targets to underpin effective therapy.

MEDUSA (Mesothelioma evolution: deciphering druggable somatic alterations) is a translational research platform to build predictive models to stratify existing therapies, while identifying new therapeutic vulnerabilities through the integration of multiomics, phenotyping and drug response data. Min Zhang, the bioinformatics director at Novogene is working with Fennell, director of the Leicester Mesothelioma Research Programme, to establish the MEDUSA platform. Large-scale gene sequencing projects, involving huge amounts of data, present new challenges in storage and analysis, making it time-consuming to detect and analyse the source and type of gene mutations. When powered by AI and deep learning, however, data patterns could emerge more effectively, such as via evolutionary clustering, a general class of problems in machine learning which applies genetic algorithms to cluster analysis.

In their pilot study, published in *Nature Communications* in 2021, Fennell and Zhang's teams integrated multiregional exome sequencing, transcriptome and multiplex immunophenotyping from 22 patients (the MEDUSA22 cohort) undergoing radical resection of mesotheliomas. Using machine learning, repeated evolutionary patterns emerged revealing conserved mutation order, with inactivation of the Hippo pathway being identified as an evolutionary bottleneck. Distinct patterns of inter-patient heterogeneity relating to prognosis and microenvironment inflammation were identified between evolutionary clusters.

"Together with Novogene, our collaborative phylogenetic analyses revealed new insights into evolutionary constraints in MPM, as well as the clinical impact of specific evolutionary trajectories," says Fennell. Their results, published in *Lancet Oncology* in 2021, also suggested that clonal architecture and evolutionary clusters dictate MPM inflammation and immune evasion which could influence the response to immune checkpoint blockade or targeted therapies.

The team is now applying integrated machine learning to decipher robust genomic features that predict response to targeted therapy, using exome sequenced MPMs from patients treated in their clinical trials, drug treated explants (from patients enrolled in MEDUSA), and primary cell lines. Response predictors will be verified in independent clinical trial cohorts led by Fennell.

"Novogene has helped optimize workflow and data integration across preclinical and clinical research to enable robust inferences, and together, we have transformed our capacity to rapidly address complex genomic questions," says Fennell. "Future studies are expected to scale our analysis by an order of magnitude and involve broader genomic and epigenomic data integration."

Zhang observed that other cancer studies can also benefit from automated bioinformatics pipelines and machine learning platforms, especially as reduced costs in sequencing technology enable wider availability of cancer genomics datasets.

"We are dedicated to equipping scientists worldwide with our gene sequencing and multiomics technologies, and addressing their preclinical, translational, clinical and diagnostic needs," he says. ∎

Novogene

Email: pr@novogne.com
Website: www.novogene.com